**Website ideas:**

“Create an analysis of existing data to make a **prediction**, classification, or regression.”  
As per final project requirements

1. Outline data set used/source:

**Credit score classification from kaggle.com**

[**https://www.kaggle.com/datasets/parisrohan/credit-score-classification**](https://www.kaggle.com/datasets/parisrohan/credit-score-classification)

A data set that contains a range of credit-related information, both numerical and categorical.

The goal of our ML model is to predict the credit score of an individual based on variables contained in the data set.

From equifax.com:

A credit score is a three-digit number designed to represent the likelihood you will pay your bills on time. There are many different credit scores and scoring models. Higher credit scores generally result in more favourable credit terms.

1. Prediction

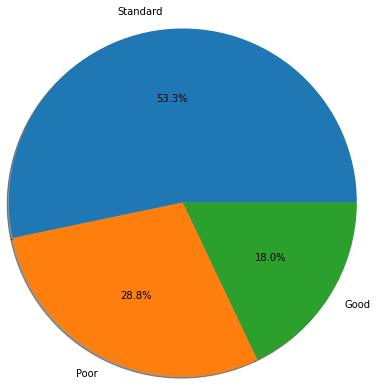
Variables contained in the data set include:

* Unnamed: 0
* Month
* age
* occupation
* annual\_income
* num\_bank\_accounts
* num\_credit\_card
* interest\_rate
* num\_of\_loan
* type\_of\_loan ...
* credit\_mix
* outstanding\_debt
* credit\_utilization\_ratio payment\_of\_min\_amount
* total\_emi\_per\_month amount\_invested\_monthly
* payment\_behaviour
* Monthly\_balance
* credit\_score
* credit\_history\_age\_new

The full definitions of these variables are available on kaggle.com

Credit score is one-hot encoded as 0,1, and 2.

The following figure represents the mix of credit score as found in the dataframe:



We assume that variables to do with monetary value such as annual\_income, outstanding\_debt, num\_of\_loan will have the highest impact on credit\_score outcome.